



Sokullu, S. (2016). A Semi-parametric Analysis of Two-Sided Markets: An Application to the Local Daily Newspapers in the USA. *Journal of Applied Econometrics*, 31(5), 843–864.  
<https://doi.org/10.1002/jae.2464>

Peer reviewed version

Link to published version (if available):  
[10.1002/jae.2464](https://doi.org/10.1002/jae.2464)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the peer reviewed version of the following article: Sokullu, S. (2015), A Semi-Parametric Analysis of Two-Sided Markets: An Application to the Local Daily Newspapers in the USA. *J. Appl. Econ.*, which has been published in final form at DOI: 10.1002/jae.2464. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# A Semi-parametric Analysis of Two-Sided Markets: An Application to the Local Daily Newspapers in the U.S.\*

Senay SOKULLU <sup>†</sup>

University of Bristol

February 4, 2015

## Abstract

This paper considers an empirical semiparametric model for two-sided markets. Contrary to existing empirical literature on two-sided markets, we specify network effects and probability distribution functions of net benefits of the two sides nonparametrically. We then estimate the model by nonparametric IV regression for local daily newspapers from the US. We show that semiparametric specification is supported by the data and the network effects are neither linear nor monotonic. With a numerical illustration we demonstrate that the mark-up of the newspaper on each side changes drastically with the nonlinearly specified network effects from the case with linear network effects.

**Keywords:** Two-sided markets, Network externality, Nonparametric IV, Ill-posed inverse problems, Tikhonov Regularization

**JEL Classification:** C14, C30, L14

---

\*I am indebted to my advisor Jean-Pierre Florens for his generous advice, support and encouragement. I thank Alexei Alexandrov, Richard Blundell, Lapo Filistrucchi, Marc Ivaldi, Gregory Jolivet, Toru Kitagawa, Costas Meghir, David Pacini, Jean-Marc Robin, Christophe Rothe, Anna Simoni, Tuba Toru, Frank Windmeijer and the seminar participants at University of Bristol and at CEMMAP for numerous helpful comments. All errors are mine.

<sup>†</sup>E-mail: *Senay.Sokullu@bristol.ac.uk*

# 1 Introduction

In the past decade a lot of work has been done both in terms of theory and in terms of empirics of two-sided markets; see, for instance, Rochet and Tirole (2003); Armstrong (2006); Rysman (2004); Kaiser and Wright (2006). However, most of these papers specify utility as a linear function of network effects and all the empirical ones use parametric specifications, see Argentesi and Filistrucchi (2007); Kaiser and Wright (2006). More precisely, the externality exerted on side one by side two is assumed to be linear in the number of agents on side two and vice versa. In this paper we develop a structural model for two-sided markets where we do not specify any functional form for the network effects and we estimate them with nonparametric instrumental variables estimation. This results in an ill-posed inverse problem which is solved using the Tikhonov Regularization scheme. The nonparametric specification of network effects allows us to capture nonlinearities and non-monotonicities in the network effect function with the increasing number of agents on the other side. The results of both nonparametric and nonlinear parametric estimation show that network effects are not linear. The implications of this result for misspecified parametric models are demonstrated with a numerical illustration.

The main feature of a two-sided market is the existence of externalities between the two sides of the market. More precisely, the benefit of agents on one side of the platform depends on the number of agents on the other side. However, as is pointed out in Rochet and Tirole (2003), a market with network externalities is a two-sided market if platforms can effectively cross-subsidize between different categories of end users. So, it is not only the interdependence of the sides to enter the platform but also the pricing structure of the platform which defines a market as two (or multi) sided. So far various industries have been examined under this setting: media, academic journals, dating agencies, credit cards, shopping malls, etc, see Rochet and Tirole (2003); Anderson and Coate (2005); Dubois et al. (2007). For example, in the newspaper industry, the decision of advertisers to advertise in a particular newspaper depends on the circulation rate of that newspaper. On the other side of the market, the readers may care about the advertising content of the newspaper they buy. The platform,

namely the newspaper, can use this interdependence between the two sides when deciding on its pricing scheme. In this case, it is going to be more aggressive with the readers if the advertisers benefit more from contacting the readers on the platform.

While in some industries, the network effect is continuously increasing, in some other industries it is nonlinear and non-monotone. For example, in the credit card industry, where the credit card is the platform and buyers and sellers are the two sides, increasing the number of sellers who accept the credit cards would unambiguously increase the benefit and thus the number of buyers who hold a credit card. However, when we think of the magazine market, although it is found that readers get utility from seeing adverts in a magazine (see Kaiser and Wright, 2006; Kaiser and Song, 2009), increasing the share of ad pages relative to content pages may start to give disutility to readers and thus make them leave the platform. In this case, the positive network externality for readers may become negative after some threshold level of ads. These network externalities play a crucial role in the platform's decision of pricing scheme since a price change for one side does not only affect the agents on that side, but it also affects the agents on the other side through the network effects. More concretely, the pricing equations of the platform can be explained by the usual *Lerner index* plus an extra term coming from the relationship with the other side of the market. An empirical study where the network effects are specified linearly may give misleading results if in fact we have nonlinear and non-monotonic network effects. In the case of nonlinear and monotonic network effects where the specification is linear, the misspecification will lead to quantitative errors, such as under or over estimation of the markup which in turn may result in erroneous conclusions about the market power of the platform. The case of nonmonotonicity can even be more crucial for two-sided markets, since the nonmonotonic network effect functions can be a source for the emergence of many platforms. An empirical study with a parametric specification of network effects may not uncover the true structure of the market and may therefore lead to erroneous conclusions in the analysis.<sup>1</sup>

In order to address this important issue, we set up a semiparametric model, in which we

---

<sup>1</sup>Ambrus and Argenziano (2009) study in detail how multiple platforms co-exist where heterogeneity in the consumer valuation of the network externality is an important ingredient.

include the network effects nonparametrically in the demand functions. This specification can show whether the increasing number of agents on one side may effect the participation decision of the agents on the other side negatively or positively. In addition to this, we do not specify any probability distribution function for the net benefits of agents, which leads to nonparametrically specified demand functions. We therefore estimate the network effect functions and the demand functions of the two sides nonparametrically. To the best of our knowledge, neither the functional specification of the network effects nor the nonparametric approach has been used in the empirical two-sided market literature before. More broadly, nonparametric approaches have not yet been used in the empirical analysis of network industries.

Nonparametric estimation has gained a lot of attention as it has many advantages. First of all, the model is not approximated by a finite set of parameters and hence not affected by any specification error. Secondly, the estimation results do not rely on parametric restrictions and show us if the economic model is really supported by the data or not. We estimate the functions of interest by nonparametric IV estimation allowing for the endogenous variables to enter the model both parametrically and nonparametrically. However, it is well known that nonparametric IV estimation causes an ill-posed inverse problem which needs to be regularized (Darolles et al., 2011; Horowitz, 2011; Newey and Powell, 2003, See). There are many papers in the literature that cope with this problem with different regularization schemes. Following the approach of Darolles et al. (2011), we regularize our inverse problem with *Tikhonov Regularization* and estimate the unknown density functions of the variables with kernels. Depending on the regularity of the function, this may give an optimal convergence rate or slower, but it does prevent the possible specification errors coming from the misspecification of the parametric form.

Although there are advantages of nonparametric specification and estimation, its use may also have limitations especially in empirical I.O., such as difficulty to make policy analysis and/or simulations in the absence of parameters. In those cases nonparametric approaches can still be used to inform the researcher on the functional forms of the model. We perform a parametric estimation whose functional specifications are based on the nonparametric estimation

results. More precisely, using the same structural equations, we approximate the unknown functions by nonlinear parametric forms, and estimate demand equations simultaneously. The results are consistent with what we have obtained in our nonparametric analysis.

Three main groups of literature are related to this paper. The first one is the two-sided markets literature. The theoretical literature on two-sided markets has focused on credit card markets, buyer-to-buyer platforms, academic journals and media. Moreover, in the theoretical two-sided markets literature the network effects are assumed to be linear most of the time, see Rochet and Tirole (2003); Armstrong (2006); Weyl (2010); Anderson and Coate (2005). The model we use in this paper is mostly related to that of Armstrong (2006), as it is more suitable to the newspaper industry, though we do allow for nonlinear and non-monotonic network effects. Empirical studies have concentrated mostly on media. Kaiser and Wright (2006) and Kaiser and Song (2009) analyze the magazine industry, whereas Argentesi and Filistrucchi (2007) and Filistrucchi and Klein (2013) are examples from the newspaper industry. Rysman (2004) uses a nonlinear specification for network effects in his paper where he examines the market for *Yellow Pages* directories. In contrast to our paper, all these papers are using parametric specifications.

There is also a vast amount of literature on the newspaper industry which started even before the two-sided markets literature. Rosse (1970) is one of the first papers who shows empirically that there are network externalities between the readers and advertisers of daily newspapers in the U.S. Thompson (1989) shows that advertising demand does not only depend on circulation but also on the reader profile of the newspaper. More recently, Fan (2013) examines how the newspaper quality, subscription prices and ad rates are affected by a change of structure in the industry. In this paper, we are also examining the local daily newspaper industry, however what we are interested in is the level of advertisements and/or readership from which each side start to get disutility or utility. With our nonparametric specification of network effects, we do not make *a priori* assumptions on the externality exerted on readers by advertisers or vice versa. We search for a threshold level from where this externality changes its sign. Our results show that the readers of the daily newspapers like advertisements, however,

their benefit starts to decrease after a certain amount of advertisements. On the other side of the market, advertisers as well get higher benefit with the higher readership and as in the case of the readers' side, their benefits start to decrease after a threshold level of circulation.

The third group of related literature is on nonparametric estimation methods with endogenous variables. Ai and Chen (2003), Newey and Powell (2003), Carrasco et al. (2007), Darolles et al. (2011), Feve and Florens (2009), Florens and Sokullu (2014) all deal with the problem of nonparametric IV estimation in the presence of endogenous variables. Although Ai and Chen (2003) and Newey and Powell (2003) use sieve methods and get over the problem of ill-posedness by putting bounds on integrals of higher order derivatives, all the other papers use kernel estimation and regularize the ill-posed inverse problem by *Tikhonov Regularization*. Florens and Sokullu (2014) is different from the other ones, in the sense that they develop their estimation technique for semiparametric transformation models. This paper also studies the nonparametric estimation of semiparametric transformation models, however, different from Florens and Sokullu (2014), all explanatory variables in our system are endogenous. Note that this extension can be relevant in many empirical models in economics as these can have more than one endogenous variables, e.g. price and market shares in nested logit models.

The paper proceeds as follows. In *Section 2*, we introduce our model, derive the structural demand equations for readers and advertisers and pricing equations of the platform. We present our semi-parametric model, its identification and estimation *Section 3*. *Section 4*, presents the empirical analysis with semi-parametric and parametric specifications while in *Section 5*, we present a numerical illustration to show the importance of misspecification of the two-sided network effects. Finally, *Section 6* concludes. All the proofs are presented in the *Online Appendix*.

## 2 The Model

In this section we introduce our model taking into account the fact that an agent on one side will consider the number of agents on the other side when she makes her decision whether to enter the platform or not. In this paper we use data from the local daily newspaper industry

to perform our empirical analyses. The two-sided market model for the newspaper industry is defined as follows: Each local daily newspaper is a monopolist in the city in which the paper is distributed. It produces content pages and ad pages for its readers and provides advertising outlet for firms who want to reach the readers of the newspaper. It maximizes its profits by setting a daily price for the readers and an advertising rate for the advertisers. Readers decide to buy the newspaper or not and firms decide to buy ad space from the newspaper or not, by looking at their net benefits. As it is a two-sided industry, the benefits of readers depend on the amount of advertisements in the newspaper and the benefits of advertisers depend on the number of readers of the newspaper, as well as some other newspaper characteristics.

We obtain the demand functions of readers and advertisers following the approach used by Larribeau (1993) and Feve et al. (2008). Let us begin with the reader side. We assume that the readers are heterogeneous in their net benefit  $b_i^r$  of buying the newspaper and these benefits are drawn from a continuous distribution. Each reader  $i$  decides to buy the local daily newspaper if her net benefit is higher than a threshold (say, her net costs)  $\underline{b}^r$ :

$$b_i^r \geq \underline{b}^r(N^a, X, U)$$

where  $\underline{b}^r(\cdot)$  is the threshold benefit level for readers which is a function of the share of advertisers on the same platform  $N^a$ , the observable newspaper characteristics  $X$  and the newspaper characteristics that are unobservable by the econometrician,  $U$ . All readers whose net benefits are higher than this threshold will buy the newspaper. Thus, the probability of buying the newspaper and hence the market share of readers is given by:

$$N^r = P(b_i^r \geq \underline{b}^r(N^a, X, U)) = 1 - F^r(\underline{b}^r(N^a, X, U)) \quad (1)$$

where  $F^r(\cdot)$  is the cdf of the net benefits of readers. We can rewrite equation (1) as:

$$N^r = S^r(\underline{b}^r(N^a, X, U)) \quad (2)$$

where  $S^r(\cdot) = 1 - F^r(\cdot)$  is the survival function. Equation (2) gives the demand of readers



for the newspaper. We assume that it is strictly decreasing in the threshold benefit level. Given the distribution of benefits of readers, the higher the threshold is, the less readers buy the newspaper. Furthermore, if the observable magazine characteristics are price and number of content pages, we expect the threshold benefit level to be increasing in cover price and decreasing in the number of content pages thus the demand is decreasing in cover price and increasing in number of content pages. The effect of the share of ad pages is ambiguous, since the readers may like the ads or not depending on the readers' tastes.

Now, let us consider the advertisers. Each advertiser  $j$  has a net benefit  $b_j^a$  from advertising in the local newspaper and these net benefits are drawn from a continuous distribution  $F^a(.)$ . They advertise in the newspaper if their net benefit is higher than  $\underline{b}^a$ , i.e:

$$b_j^a \geq \underline{b}^a(N^r, W, V)$$

where  $\underline{b}^a(.)$  is the threshold benefit level for advertisers and it is a function of the share of readers of the newspaper  $N^r$ , observable newspaper characteristics for advertisers  $W$  and the newspaper characteristics that are unobservable by the econometrician,  $V$ . Like in the case of readers, the probability of advertising in the newspaper and thus the share of advertisers who join the newspaper is given by<sup>2</sup>:

$$N^a = P(b_j^a \geq \underline{b}^a(N^r, W, V)) = S^a(\underline{b}^a(N^r, W, V)) \quad (3)$$

where  $S^a(.) = 1 - F^a(.)$ . Equation (3) is the demand equation of advertisers for the local newspaper. It is strictly decreasing in the threshold benefit function. The threshold benefit function is expected to be increasing in the ad rate and decreasing in the share of readers. So, more firms would like to advertise in a newspaper with a higher readership and a lower ad rate.

---

<sup>2</sup>In the application we define the market size for the advertising side as the total number of pages in the monopolist local daily newspaper, hence the share of advertisers who join the platform will be equal to the share of ad pages in the newspaper.

Now, we can write the demand system for the newspaper:

$$N^r = S^r(\underline{b}^r(N^a, X, U)) \quad (4)$$

$$N^a = S^a(\underline{b}^a(N^r, W, V)) \quad (5)$$

Given the demand equations of both sides, the monopolist local newspaper chooses its cover price and advertising rate to maximize its profit:

$$\max_{P^r, P^a} \Pi = \max_{P^r, P^a} \{P^r N^r M^r + P^a N^a M^a - C(N^r M^r, N^a M^a, IP) - FC\} \quad (6)$$

where  $P^{(i)}$  and  $M^{(i)}$ ,  $i = a, r$  are price and market size for both sides,  $C$  is the cost function which depends on the number of agents on each side as well as other cost variables such as input prices,  $IP$ , and  $FC$  is the fixed costs. The maximization problem in equation (6) gives the following pricing equations:

$$\frac{P^r - \frac{\partial C}{\partial N^r}}{P^r} = -\frac{1}{\epsilon_{N^r}} - \frac{\left(P^a - \frac{\partial C}{\partial N^a}\right) \frac{\partial N^a}{\partial P^r}}{\frac{\partial N^r / \partial P^r}{P^r}} \quad (7)$$

$$\frac{P^a - \frac{\partial C}{\partial N^a}}{P^a} = -\frac{1}{\epsilon_{N^a}} - \frac{\left(P^r - \frac{\partial C}{\partial N^r}\right) \frac{\partial N^r}{\partial P^a}}{\frac{\partial N^a / \partial P^a}{P^a}} \quad (8)$$

where  $\epsilon_{N^r}$  and  $\epsilon_{N^a}$  are price elasticity of reader demand and price elasticity of advertiser demand, respectively. It should be noted that both equations (7) and (8) are modified versions of the *Lerner Index*, in the sense that they also include the network externalities coming from the two-sidedness of the industry. For example, for the reader side, the mark-up (the term on the left hand side), which is the ability of the newspaper to price over its marginal cost, depends on the inverse price elasticity of the readers,  $1/\epsilon_{N^r}$  as well as on a second term. This second term captures the externality that readers have on advertisers. If they exert a positive externality on advertisers, the newspaper charges readers a lower price compared to a situation where this externality is ignored. The intuition is simple. By lowering the cover price, the newspaper can attract more readers which in turn attract more advertisers through

network externalities, thus increasing the profits of the newspaper. Put differently, as Weyl (2010) shows, the newspaper indeed still picks the price where marginal revenue is equal to marginal cost. However, in this special case, the marginal revenue and marginal cost have extra terms coming from the network externalities between the two sides.

It is therefore important to identify the sign and magnitude of the externalities as they play a crucial role in pricing. In the next section, where we conduct an empirical analysis of the industry, we specify the network effects as unknown functions to be able to see if the exerted externality changes sign with an increasing number of ad pages or not.

### 3 A Semiparametric Model and Estimation

In this section, we introduce the model and the nonparametric IV estimator we use to do the empirical analysis of the local daily newspaper industry.

#### 3.1 Model Specification

The daily newspaper industry has been studied extensively since the 1960's. Rosse (1970), which is one of the most influential papers, shows empirically that circulation and advertising depends positively on each other in the U.S. daily newspaper industry. Ferguson (1983) then supports this idea by finding that circulation demand depends not only on quantity but also on the value of information provided by retail advertising. Nonetheless, when we consider more recent literature, Argentesi and Filistrucchi (2007) do not find a significant effect of advertisements on reader demand for Italian newspapers. Neither does Fan (2013) who studies daily newspapers in the U.S. On the other side of the market, the dependence of the advertising demand on circulation has always been found to be positive, although Thompson (1989) shows that the demand for display advertisement can be sensitive to the paper's readership profile.<sup>3</sup> In this section our aim is to see if these network externalities between the two sides change sign as the share of agents on each side increases, rather than to

---

<sup>3</sup>There are two main types of advertising: Display advertising and classified advertising. Display advertising can appear throughout the paper and usually involves illustrations. Classified advertising mostly appears on special pages under the relevant heading for the advertised items, see Ferguson (1983).

see if it is negative or positive. So, instead of making a linear parametric specification for two sided network effects, we specify them with nonparametric functions to be able to capture the variation in network externality with the variation in the share of agents on the other side.

Evans and Schmalensee (2008) and Argentesi and Ivaldi (2005) point out that failure to account for network externalities in two-sided platforms can lead to serious errors in antitrust analysis. Correct specification of these network effects is important for the same reason. However, none of the aforementioned papers on the daily newspaper industry allows for nonlinear and nonmonotone network effects in the utility function of agents. It is straightforward to see e.g. in case of nonlinearity or non-monotonicity of the network effect functions, the results of Kaiser and Song (2009) regarding the elasticities would not be the same.<sup>4</sup> Rysman (2004) adopts a nonlinear specification for network effect functions. However, his specification does not allow for non-monotonicity. It can be easily shown that, if non-monotonicities exist in the network effect functions, the market equilibrium condition he derives would be harder to satisfy, i.e. it may not be satisfied for all values of the advertisements. Thus, his results may not hold anymore if the network effect of advertisers on consumers are nonmonotone.

In this semiparametric model specification, we make as few parametric approximations as we can. First of all, we make no assumption on the family of distribution functions of net benefits of readers and advertisers. Secondly, we assume that network effects are given by some unknown functions,  $\varphi(N^a)$  and  $\psi(N^r)$ . Finally we specify the threshold benefit functions  $\underline{b}^r$  and  $\underline{b}^a$  as linear functions of network externalities and platform characteristics. Then, the system of demand equations are given by the following:

$$N^r = S^r(\varphi(N^a) + X\beta + U) \quad (9)$$

$$N^a = S^a(\psi(N^r) + W\gamma + V) \quad (10)$$

---

<sup>4</sup>Suppose that the network effects are specified as second order polynomials in Kaiser and Song (2009), so that the mean utility is given by:  $\delta_{jt} = X_{jt}\beta + \theta_0 + \theta_1 Ad_{jt} + \theta_2 Ad_{jt}^2 - \alpha p_{jt} + \gamma_t + \eta_j + \xi_{jt}$  instead of  $\delta_{jt} = X_{jt}\beta + \theta Ad_{jt} - \alpha p_{jt} + \gamma_t + \eta_j + \xi_{jt}$ . In the first case the advertising elasticity of demand will be given by  $(\theta_1 + 2\theta_2 Ad_{jt})s_{jt}(1 - s_{jt})$  whereas in the second case it will be given by  $\theta s_{jt}(1 - s_{jt})$ , where  $s_{jt}$  is the market share of magazine  $j$  on the readers' side. Then for example, if  $\theta_2$  is estimated to be negative and  $\|\theta_1\| < \|2\theta_2 Ad_{jt}\|$  the elasticity would be negative compared to a case where  $\theta$  is estimated to be positive.

We are interested in estimating the network externalities between the two sides, hence for simplicity, we consider just one platform characteristic, price, whose coefficient is normalized to one for identification. The reason to use just one characteristic is that since we use kernels in estimation, increasing the dimension of endogenous and/or exogenous variables complicates the estimation process. Note that this way of specification does not allow us to analyze the elasticities however in this paper we are interested in identifying the shape of the network effect functions and we can make this normalization without loss of generality. This final specification gives us the equations we are going to estimate:

$$N^r = S^r(\varphi(N^a) + P^r + U) \quad (11)$$

$$N^a = S^a(\psi(N^r) + P^a + V) \quad (12)$$

where  $P^r$  is the daily price of the newspaper and  $P^a$  is the ad rate per column-inch.

We will estimate the functions of interest, namely,  $S^r$ ,  $S^a$ ,  $\varphi$  and  $\psi$  nonparametrically.

### 3.2 Identification

In this section we closely follow Florens and Sokullu (2014). The assumptions needed for identification of nonparametric IV models are quite standard, see Darolles et al. (2011); Newey and Powell (2003); Florens and Sokullu (2014); Berry and Haile (2014). Before proceeding to discuss identification, we introduce our variables and some notation.

$N^r$ ,  $N^a \in \mathbb{R}$  are the endogenous market shares of the newspaper on the readers' and advertisers' side, respectively. The platform characteristics for readers and advertisers, which we denote by  $X$  and  $W$  in equations (9) and (10), can be endogenous or exogenous. These characteristics generally include prices on both sides, newshole, number of reporters, edition dummy indicating if it is a morning or evening paper, etc. (see Fan, 2013). As already mentioned, for simplicity, we include just one platform characteristic in each function, namely daily price ( $P^r$ ) and ad rate ( $P^a$ ). Thus,  $P^r, P^a \in \mathbb{R}$ , are endogenous. Let  $Z^r \in \mathbb{R}^{q^r}$  and

$Z^a \in \mathbb{R}^{q^a}$  denote instruments for each equation and  $Z = \{Z^r, Z^a\}$ .<sup>5</sup> Finally, unobservable newspaper characteristics for each side,  $U$  and  $V$  are scalars as well.

We assume that  $(N^r, N^a, P^r, P^a, Z)$  generate a random vector,  $\Xi$ , whose density is square integrable and which has a cumulative distribution function  $F$ . Then for each  $F$ , we can define subspaces of our variables as  $L_F^2(N^r), L_F^2(N^a), L_F^2(P^r), L_F^2(P^a)$  and  $L_F^2(Z)$  which belong to a common Hilbert space. That is,  $L_F^2(Z)$  denotes the subspace of  $L_F^2$  of real valued functions depending on  $Z$  only. In the sequel, we use the notation  $L_Z^2$  to denote the  $L_F^2(Z)$ .

Now we can state the needed assumptions for identification.

**Assumption 1 *Strict monotonicity.*** *The survival function  $S^q(d)$ ,  $q \in \{r, a\}$  is strictly decreasing in  $d$ .*

As  $S$  is a survival function, we know that it is decreasing, however, by making the assumption of "strictly decreasing", we guarantee to have a unique inverse of it, which we will use for identification and estimation. Thus, using Assumption 1, we can rewrite the system of equations in (11) and (12) as:

$$H^r(N^r) = \varphi(N^a) + P^r + U \quad (13)$$

$$H^a(N^a) = \psi(N^r) + P^a + V \quad (14)$$

where  $H^q(N^q) = (S^q)^{-1}(N^q)$  for  $q = a, r$ .

**Assumption 2 *Conditional mean independence.***  $\mathbb{E}[U|Z^r] = 0$  and  $\mathbb{E}[V|Z^a] = 0$

**Assumption 3 *Completeness.***  $(N^r, N^a)$  are strongly identified by  $Z^q$  for  $q = a, r$ :

$$\forall m(N^r, N^a) \in L_{N^r}^2 \times L_{N^a}^2, \quad \mathbb{E}[m(N^r, N^a)|Z^q] = 0 \quad \Rightarrow \quad m(N^r, N^a) = 0 \quad a.s \quad for \quad q \in \{a, r\}$$

**Assumption 4 *Measurable separability.***  $N^r$  and  $N^a$  are measurably separable:

$$\forall m \in L_{N^r}^2, l \in L_{N^a}^2, m(N^r) = l(N^a) \Rightarrow m(.) = l(.) = constant$$

---

<sup>5</sup>The instruments will be introduced in the application.

**Assumption 5 Normalization.**

$$\forall l \in L_{N^a}^2, l(.) = \text{constant} \Rightarrow \text{constant} = 0$$

$$\forall m \in L_{N^r}^2, m(.) = \text{constant} \Rightarrow \text{constant} = 0$$

For simplicity, we will assume that  $\varphi(.)$  and  $\psi(.)$  are normalized by the conditions  $\mathbb{E}(\varphi(N^a)) = 0$  and  $\mathbb{E}(\psi(N^r)) = 0$ . Under this assumption, the parametric space we consider is:

$$\mathcal{E}_0^q = \{(H^q, \phi) \in L_{N^r}^2 \times L_{N^a}^2 \text{ such that } \mathbb{E}[\phi] = 0, q = a, r\}$$

*Assumption 2* is just a conditional mean independence condition. *Assumption 3, Completeness* is the nonparametric counterpart of the rank condition in parametric IV estimation and it is also referred to as *complete statistic* in the statistics literature (See Lehmann and Scheffe, 1950; Basu, 1955). It is a condition on the power of  $Z^r$  and  $Z^a$  to identify  $H^r(.)$  and  $\varphi(.)$  and  $H^a(.)$  and  $\psi(.)$ , respectively. Intuitively, it means that there is no function of  $N^r$  and  $N^a$  that is not correlated with any function of  $Z^r$  and  $Z^a$ . The completeness assumption is an assumption on the distribution of the endogenous variables conditionally on the instruments. Although, in this paper we take it as given, further reference on the primitive conditions for completeness can be found in D' Haultfoeuille (2011), Andrews (2011) and Hu and Shiu (2011). As it is shown by Hu and Shiu (2011), our completeness condition on the multivariate distribution function of the endogenous variables can be obtained by the completeness on univariate distribution functions and thus it is not very restrictive. *Assumption 4, Measurable separability*, is needed to distinguish  $H^r(.)$  and  $\varphi(.)$  and  $H^a(.)$  and  $\psi(.)$ . It says that for all values of  $N^r$  and  $N^a$ , two functions  $m(N^r)$  and  $l(N^a)$  can be equal only if they are equal to a constant. In other words, it states that there is no exact relation between  $N^r$  and  $N^a$ . So, for the reader demand equation, *Assumption 4* holds whenever  $P^r + U$  has some components that vary independently of  $N^a$  and for the advertiser demand equation, it holds whenever  $P^a + V$  has some components that vary independently of  $N^r$ . It is not hard to verify measurable separability in our model. In the model, we already state that reader demand depends on the

share of advertisers  $N^a$ , daily price  $P^r$ , as well as some unobservable newspaper characteristics for readers,  $U$ . Hence, we do not make an unrealistic assumption by measurable separability, as it is natural to expect that  $U$  includes some newspaper characteristics for readers that vary independently of the share of advertisers. Moreover, with the same reasoning, it is natural to expect that there exists some unobservable newspaper characteristics for advertisers that vary independently of the share of readers. Finally, *Assumption 5* is just a normalization assumption for identification.

**Theorem 1** *Under the assumptions 1-5, the functions  $S^r$ ,  $S^a$ ,  $\varphi$  and  $\psi$  are identified.*

In this paper we are considering a fully recursive system of equations. For this reason we now state extra assumptions that guarantee the existence of the reduced form solution of the system. In other words, we show that  $N^r$  and  $N^a$  can be explained by the other variables of the system for given cover price and ad rate. Blundell and Matzkin (2010) also discuss the existence of a reduced form solution to the system of equations where they investigate the control function approach in nonparametric nonseparable simultaneous equations models.

Remember that our structural system is given by equations (11) and (12).

We further make the following assumptions:

**Assumption 6** *For all values of  $N^r, N^a, P^r, P^a, U$  and  $V$ , the functions  $S^r$  and  $S^a$  are continuously differentiable.*

This assumption requires that the net benefits of readers and advertisers are distributed continuously so that the survival functions will be continuously differentiable on the interval  $(0, 1)$ .

**Assumption 7** *For all values of  $N^r, N^a, P^r, P^a, U$  and  $V$ , we have:  $\frac{\partial S^r}{\partial N^a} \frac{\partial S^a}{\partial N^r} < 1$*

This assumption is indeed a restriction on the magnitude of the network effects. Note that similar types of assumptions are made by Blundell and Matzkin (2010) and Filistrucchi and Klein (2013). Blundell and Matzkin (2010) state that this assumption is to relax the common



condition of the intersection of an upward and a downward sloping functions to determine the value of the endogenous variables, which are in our case reader and advertiser demands.

Now, we can state the theorem for the existence of a reduced form solution:

**Theorem 2** *Under assumptions 1,6,7 and for given cover price and ad rate, there exist unique functions  $h^r$  and  $h^a$  representing the structural model in equations (11) and (12), such that:*

$$N^r = h^r(P^r, P^a, U, V) \quad \text{and} \quad N^a = h^a(P^r, P^a, U, V)$$

The result of Theorem 2 has important implications in terms of the existence of a unique equilibrium. For the given prices, the existence of unique functions for  $N^r$  and  $N^a$  which don't depend on each other, indeed shows that there exists a unique equilibrium in the coordination game between the two sides of the market. This is an important result since it has already been mentioned in the literature that demand may not be unique for given prices due to this coordination problem (see Rochet and Tirole, 2003; Caillaud and Jullien, 2003; Armstrong, 2006). Filistrucchi and Klein (2013) adopted a similar approach to ours to show that unique reduced form demand equations exist for given prices. Given these reduced form demand functions, one can rewrite the profit function of the newspaper and then derive the conditions that guarantee the concavity of the profit function. We leave the derivation of these conditions for future work since it is beyond the scope of this paper.

**Remark 3** *In this section we present the identification of a separable model since we cannot identify the network effect functions separately in a nonseparable model. Nonetheless, we present our model under a nonseparable framework and discuss its identification and estimation in Online Appendix A.*

### 3.3 Nonparametric Estimation of Semiparametric Transformation Equations

We estimate equations (13) and (14) with the nonparametric instrumental variable estimator for semiparametric transformation models. As already stated, nonparametric IV regression

has gained a lot of attention recently. The model we present in this section is different from those found in the existing literature in the sense that it covers a very general case. We allow for nonparametric specifications on the right and left hand side of the transformation model, as well as for endogeneity, and use the mean independence condition rather than full independence. Although Horowitz (1996) and Linton et al. (2008) study the nonparametric estimation of transformation models, none of them consider the case of nonparametric specifications on both sides of the regression equation. Our method is very similar to that of Florens and Sokullu (2014), since both are nonparametric estimators of semiparametric transformation models which allow for nonparametric specifications on both sides of the equation. Florens and Sokullu (2014) consider a partially linear equation and assume that the parametric component is exogenous, i.e, they consider the model  $H(Y) = \varphi(Z) + X + U$  where  $\mathbb{E}(U|Z) \neq 0$  and  $\mathbb{E}(U|X) = 0$ . In this paper, we consider the model given in Equations (13) and (14) where  $\mathbb{E}(U|N^a) \neq 0$ ;  $\mathbb{E}(U|P^r) \neq 0$ ;  $\mathbb{E}(V|N^r) \neq 0$  and  $\mathbb{E}(V|P^a) \neq 0$ . More precisely, the parametric component is no longer assumed to be exogenous. Given this difference, the convergence rates of the operators as well as the proof of consistency are adapted to our case. Different from Florens and Sokullu (2014), we show in detail how to implement the estimator, though it should be noted that the implementation of a nonparametric IV estimator is now commonplace in the literature (see Darolles et al., 2011; Florens et al., 2009; Fève and Florens, 2009). Moreover, we show that under the assumption of a monotonic transformation the estimation of  $H^r(\cdot)$  and  $H^a(\cdot)$  can be improved by monotonisation such as rearrangement. We further obtain bootstrap confidence intervals for a fully simultaneous system.

We follow a limited information approach while presenting nonparametric IV estimation for transformation equations. In other words, we estimate the demand system given by equations (13) and (14) equation by equation. For the rest of this section, we continue to present the estimation method using equation (13) only. Extension to equation (14) is straightforward. Estimation follows closely Florens and Sokullu (2014) *Section 2.2* but here we provide detailed discussion of the implementation of the method.

By Assumptions 1 and 2, we can write our estimation problem as:

$$\mathbb{E}[H^r(N^r) - \varphi(N^a) - P^r|Z^r] = 0 \quad (15)$$

From now on, for ease of presentation we will work with operators. Let us define the following operator:

$$T^r : \mathcal{E}^r = \left\{ L_{N^r}^2 \times \tilde{L}_{N^a}^2 \right\} \mapsto L_{Z^r}^2 : T^r(H^r, \varphi) = \mathbb{E}[H^r - \varphi|Z^r]$$

where  $\tilde{L}_{N^a}^2 = \{ \varphi \in L_{N^a}^2 : \mathbb{E}(\varphi) = 0 \}$ . We use this projected space in order to satisfy the normalization assumption for identification, *Assumption 5*. Without this constraint on the space we cannot identify the function  $\varphi$ . The inner product is defined as:

$$\langle (H_1^r, \varphi_1), (H_2^r, \varphi_2) \rangle = \langle H_1^r, H_2^r \rangle + \langle \varphi_1, \varphi_2 \rangle$$

The adjoint operator of  $T^r$ ,  $T^{r*}$  satisfies:

$$\langle T^r(H^r, \varphi), \xi \rangle = \langle (H^r, \varphi), T^{r*}\xi \rangle$$

for any  $(H^r, \varphi) \in \mathcal{E}^r$  and  $\xi \in L_{Z^r}^2$ . From the equality above it follows immediately that

$$T^{r*}\xi = (\mathbb{E}[\xi|N^r], -\mathbb{P}\mathbb{E}[\xi|N^a])$$

where  $\mathbb{P}$  is the projection operator which project the functions of  $N^a$  on the space where those functions have zero mean. Formally:  $\mathbb{P} : L_{N^a}^2 \mapsto \tilde{L}_{N^a}^2 : \mathbb{P}\varphi = \varphi - \mathbb{E}(\varphi)$ . Since we assume that the density of  $\Xi$  is square integrable, the operator  $T^r$ , its adjoint operator,  $T^{r*}$  and the two self adjoint operators  $T^{r*}T^r$  and  $T^rT^{r*}$  are all compact, see Darolles et al. (2011).

Now we can rewrite our estimation problem using the operator notation:

$$T^r(H^r(N^r), \varphi(N^a)) = f^r \quad (16)$$

where  $f^r = \mathbb{E}(P^r|Z^r)$ .<sup>6</sup> This equation is a *Fredholm integral equation of the first kind* and the solution needs the inversion of the operator  $T^r$ . Since  $T^r$  is a compact operator its inverse is not continuous which will lead the solution not to be continuous either. This discontinuity causes an ill-posed inverse problem. In other words, this equation violates one of the definitions of the well-posedness of a problem.<sup>7</sup> More precisely, as  $T^r$  is a compact linear operator, it has infinitely many eigenvalues in the neighbourhood of zero, which makes the inverse of it,  $(T^r)^{-1}$  discontinuous.<sup>8</sup> As a result, a very small change in the value of  $f^r$  may lead the solution to explode. Intuitively, in finite dimensional case, this problem amounts to having a matrix  $M$  with zero eigenvalues and thus noninvertable.<sup>9</sup>

It should be noted that the reason of ill-posedness in this problem is the endogeneity of the market share of the advertisers. If our model was instead  $Y = \varphi(X) + U$  with  $\mathbb{E}(U|X) = 0$ , the solution to  $\varphi(X)$  will be given by  $\mathbb{E}(Y|X) = \varphi(X)$ , which would not necessitate the inversion of any operator. However, the case where  $\mathbb{E}(U|X) \neq 0$  the estimation of  $\varphi$  is an ill-posed inverse problem.<sup>10</sup> Equivalently, if we had endogeneity and we were doing parametric IV, this would not cause an ill-posed inverse problem either since the operator we were using would be a finite dimensional continuous operator whose inverse exists and is continuous.

To solve this ill-posed problem, we need to regularize it, i.e., we need to modify the operator such that the solution is not unstable, and such that this amount of modification approaches to zero as the sample size increases. For this, we choose the *Tikhonov Regularization* scheme. Under this regularization scheme, the norm of the solution is controlled by a penalty term,  $\gamma$ , which is called *regularization parameter*. The ill-posed inverse problem literature offers other regularization schemes, as well (see Carrasco et al., 2007).

The regularized solution to the identifying relation in (16) is given by the following mini-

---

<sup>6</sup>Note that in the case of Florens and Sokullu (2014), we would have an exogenous variable  $X$  instead of the  $P^r$  and then the  $f^r$  would be equal to  $\mathbb{E}(X|X, Z^r) = X$ .

<sup>7</sup>As defined in Engl et al. (1996), a problem is well-posed if the definitions below hold:

- (i) For all admissible data a solution exist.
- (ii) For all admissible data the solution is unique.
- (iii) The solution continuously depends on the data.

<sup>8</sup>See Theorem 7.22 and Theorem 7.23 in Ryanne and Youngson (2008).

<sup>9</sup>For a more detailed discussion see Horowitz (2011) which explains the ill-posedness using singular value decomposition of the operators and demonstrates it with an example.

<sup>10</sup>This problem has been studied extensively by Darolles et al. (2011); Newey and Powell (2003); Hall and Horowitz (2005).

mization program:

$$\min_{H^r, \varphi} \left\{ \|T^r(H^r, \varphi) - f^r\|^2 + \gamma_n^r \|(H^r, \varphi)\|^2 \right\}$$

where  $\gamma_n^r > 0$  and  $\gamma_n^r$  converges to zero at a suitable rate. Hence,

$$(H^r(N^r), \varphi(N^a))' = (\gamma_n^r I + T^{r*} T^r)^{-1} T^{r*} f^r \quad (17)$$

where  $I$  is the identity operator in  $L_{N^r}^2 \times L_{N^a}^2$ . We can write the solution in (17) as follows:

$$(\gamma_n^r I + T^{r*} T^r)(H^r(N^r), \phi(N^a)) = T^{r*} f^r \quad (18)$$

Equivalently:

$$\begin{pmatrix} \gamma_n^r H^r + \mathbb{E}[\mathbb{E}(H^r|Z^r)|N^r] - \mathbb{E}[\mathbb{E}(\varphi|Z^r)|N^r] \\ \gamma_n^r \varphi - \mathbb{P}\mathbb{E}[\mathbb{E}(H^r|Z^r)|N^a] + \mathbb{P}\mathbb{E}[\mathbb{E}(\varphi|Z^r)|N^a] \end{pmatrix} = \begin{pmatrix} \mathbb{E}[\mathbb{E}(P^r|Z^r)|N^r] \\ -\mathbb{P}\mathbb{E}[\mathbb{E}(P^r|Z^r)|N^a] \end{pmatrix} \quad (19)$$

To explain the implementation of the defined method, suppose that we have an i.i.d. sample of  $(N_i^r, N_i^a, P_i^r, P_i^a, Z_i)$ ,  $i = 1, \dots, n$ . As we do not know the true distribution of our variables, we need to replace the conditional expectations with their empirical counterparts.

Hence, we rewrite the system of equations with kernels.<sup>11</sup> For the system in (19), let  $A_{Z^r}$  be the matrix whose (i,j)th element is:

$$A_{Z^r}(i, j) = \frac{K_{Z^r, h_{Z^r}}(Z_i^r - Z_j^r)}{\sum_j K_{Z^r, h_{Z^r}}(Z_i^r - Z_j^r)} \quad (20)$$

Let  $A_{N^r}$  and  $A_{N^a}$  be the matrices with the (i,j)th elements:

$$A_{N^r}(i, j) = \frac{K_{N^r, h_{N^r}}(N_i^r - N_j^r)}{\sum_j K_{N^r, h_{N^r}}(N_i^r - N_j^r)} \quad (21)$$

---

<sup>11</sup>The definition of the multivariate kernel used in implementation can be found in Florens et al. (2009).

$$A_{N^a}(i, j) = \frac{K_{N^a, h_{N^a}}(N_i^a - N_j^a)}{\sum_j K_{N^a, h_{N^a}}(N_i^a - N_j^a)} \quad (22)$$

for some bandwidth parameters  $h_{N^r}$ ,  $h_{N^a}$  and  $h_{Z^r}$ . Moreover let  $A_p$  be the matrix with  $\frac{n-1}{n}$  on the diagonal and  $-\frac{1}{n}$  elsewhere. We can rewrite the system in (19) as:

$$\begin{pmatrix} \gamma_n^r \hat{H}^r + A_{N^r} A_{Z^r} \hat{H}^r - A_{N^r} A_{Z^r} \hat{\varphi} \\ \gamma_n^r \hat{\varphi} - A_p A_{N^a} A_{Z^r} \hat{H}^r + A_p A_{N^a} A_{Z^r} \hat{\varphi} \end{pmatrix} = \begin{pmatrix} A_{N^r} A_{Z^r} P^r \\ -A_p A_{N^a} A_{Z^r} P^r \end{pmatrix} \quad (23)$$

Then the estimated functions are given by:

$$\begin{pmatrix} \hat{H}^r \\ \hat{\varphi} \end{pmatrix} = \begin{pmatrix} \gamma_n^r I + A_{N^r} A_{Z^r} & -A_{N^r} A_{Z^r} \\ -A_p A_{N^a} A_{Z^r} & \gamma_n^r I + A_p A_{N^a} A_{Z^r} \end{pmatrix}^{-1} \begin{pmatrix} A_{N^r} A_{Z^r} P^r \\ -A_p A_{N^a} A_{Z^r} P^r \end{pmatrix} \quad (24)$$

Equation (24) is a system of  $2n$  equations in  $2n$  unknowns which means that we can recover  $\hat{H}^r$  and  $\hat{\varphi}$ , hence  $\hat{S}^r$ . In the *Online Appendix C*, we prove that the estimator is consistent and we compute its rate of convergence.

## 4 Empirical Analysis of the U.S. Daily Newspaper Industry

In this section we make an application of the nonparametric estimation defined in the previous section. Using the specification in *Section 3.1* and data on local daily newspapers in the U.S., we estimate the network effect functions nonparametrically.

### 4.1 Data

We have a two-sided market model of a monopoly platform, the daily newspaper industry in the U.S. High fixed costs are one of the main characteristics of this industry which leads to monopolist newspapers in most of the markets (cities), see *Editor & Publisher International Databook (2012)*, Rosse (1970); Blair and Romano (1993).<sup>12</sup>

---

<sup>12</sup>In the previous version of this paper the application is done by using a dataset on German magazine industry which was an oligopoly. We thank the anonymous referee for pointing out that the daily newspaper market data from the U.S. would fit our application better.

Our data comes from three sources: *Editor & Publisher International Databook (2012)*, *Kantar Media Intelligence* and *US Census Bureau*. Daily average circulation, subscription prices, frequency of edition, number of special editions and special weekly selections, average number of pages and other newspaper characteristics are taken from *Editor & Publisher International Databook (2012)*. We get the information on advertising quantity in column-inches and adrate per column-inch from *Kantar Media Intelligence*. Finally demographic characteristics of cities are obtained from *US Census Bureau*. Although we have more than 1000 observations on circulation, daily price and newspaper characteristics, after matching it with the advertising data we end up with a sample of 117 newspapers.

Table 1 present the summary statistics of our main variables. Advertising quantity is measured in column-inches. The mean adrate in our sample is \$62.44 per column-inch while maximum and minimum rates are highly different from this mean value at \$8.71 and \$238.9, respectively. Mean daily subscription price is \$0.54 with a minimum of \$0.17 and a maximum of \$1.31. In the data, the circulation rate is given as the daily average over the six months period covering April 1, 2011 to September 30, 2011. The average weekday circulation in our data is 97769 and average advertising quantity per daily issue is 2473 column-inches which is approximately 19.5 pages.<sup>13</sup> Average total pages per issue is 57.5 which implies on average about one third of the pages of a daily newspaper in the U.S. goes to advertising.

Table 1: Summary Statistics of the data

Variable	Mean	Std Dev.	Min	Max
Circulation	97769.94	91077.02	5338	527568
$N^r$	$5.69 \times 10^{-3}$	$6.13 \times 10^{-3}$	$0.76 \times 10^{-3}$	$3.68 \times 10^{-2}$
Daily price	0.54	0.174	0.254	1.31
Advertising (c-inches)	2473.08	1387.85	23.46	7219.88
$N^a$	0.36	0.17	0.003	0.85
Adrate	62.44	47.68	8.71	238.90
Page number per issue	57.56	27.63	13	240

We define the market size for the readers as the population of the city. However, when we construct the readers' share variable we obtain numbers greater than 1. To deal with this issue we change the definition of the market size, instead we assume that it is proportional

---

<sup>13</sup>Fan (2013) points out that a typical US daily newspaper page has 6 columns with 21 inch depth, so average number of ad pages is given by: advertising quantity in column-inches/126.

to the population by a factor 100. On the advertisers' side the market size is assumed to be all pages in the newspaper, thus the advertisers' share is obtained by dividing advertising quantity by the total column inches in the newspaper.

**Remark 4** *We have a sample with 117 observations to estimate our model nonparametrically. In Online Appendix D, we present a simulation to show that the estimation method we propose performs well with a sample of 100 observations.*

## 4.2 Semiparametric Estimation

In this section we estimate the network externality functions  $\varphi(\cdot)$  and  $\psi(\cdot)$  as well as the inverse survival functions  $H^r(\cdot)$  and  $H^a(\cdot)$  given in equations (13) and (14).<sup>14</sup> We adopt the estimation technique defined in *Section 3.3*.

All the explanatory variables in equations (13) and (14) are endogenous, so we need to use instruments. Since  $U$  and  $V$  are unobserved newspaper characteristics, we choose our instruments from city level demographic variables, see Rosse (1970); Ferguson (1983); Fan (2013). We instrument daily price with city area in square miles. Rosse (1970) as well as Ferguson (1983) mention that the city area in square miles can reflect higher distribution costs. Hence as a cost side variable it could be correlated with price but should be independent of the unobserved newspaper characteristics in the reader demand equation. Moreover, as pointed out in George (2007) local retail advertising makes up 45% of the total newspaper advertising which makes city level retail sales a good candidate to instrument advertising share in the demand equation of the readers.<sup>15</sup> We use wages in the printing industry at the county level as an instrument for ad rate. Wages in printing industry would affect the ad rate in the newspaper since it affects the cost, but it should be independent of the unobserved newspaper characteristics for the advertising side. Finally to instrument the readers' share in the advertising demand we use the city level percentage of population below the poverty level.

---

<sup>14</sup>It should be noted that our main object is to estimate network effects on each side nonparametrically. For this reason, we give our attention to the functions of interest,  $\varphi$  and  $\psi$ , and leave a more complex analysis for future work.

<sup>15</sup>The relation between advertising demand and local retail sales is also established by Rosse (1970) and Ferguson (1983).



Since this is an income related variable it will be correlated with the demand of readers for the newspaper but independent of the unobservable newspaper characteristics in the advertising demand equation. Correlation of income and/or income related variables with circulation has already been established by Ferguson (1983) and Fan (2013). We check the correlation of the instruments with the endogenous variables. The correlation coefficient between the cover price and city area in square miles is equal to 0.006 while the correlation coefficient between the advertising share and city level retail sales is equal to  $-0.22$ . For the advertising demand equation, the correlation coefficient of price and readers' share with the corresponding instruments are equal to 0.01 and  $-0.15$ , respectively.

We use a *rule of thumb* to construct the bandwidth parameters. The regularization parameters  $\gamma_n^r$  and  $\gamma_n^a$  are chosen by the data-based selection rule of regularization parameter proposed in Florens and Sokullu (2014). Furthermore, since the survival functions are monotonic, we monotonized these functions ( $H^r$  and  $H^a$ ) both by *isotonisation* and by *rearrangement*, after the estimation.<sup>16</sup> We report the results where the monotonization is done by rearrangement since it gives better results for the monotonization of probability distribution functions compared to other existing methods, as is pointed out by Chernozhukov et al. (2010). Finally, we construct pointwise bootstrap confidence intervals for the estimated functions. We obtain the confidence interval with 500 replications of *bootstrap in pairs*.<sup>17</sup> In estimation with the bootstrapped data, the values of the regularization and bandwidth parameters are fixed at the levels that we use in the original estimation.

The results are given in Figures 1 and 2.<sup>18</sup>

The estimated network externality functions do not exhibit the same pattern. The network effect of advertisers,  $\varphi$  is decreasing up to 40% of the ad share. Up to 30% of ad share it takes positive values, and after that point, it takes negative values. In our set up, the share of agents joining the platform is given by the survival function of their net benefits and the survival function is decreasing in its arguments. So, it means that, up to an advertiser share

---

<sup>16</sup>For *isotonisation* and *rearrangement*, see Chernozhukov et al. (2009).

<sup>17</sup>Bootstrap in pairs is performed by resampling the data directly by replacement. We do bootstrap in pairs since all the explanatory variables in the model are endogenous. (See Horowitz, 2001; Flachaire, 2005)

<sup>18</sup>We also present estimation results without monotonization and pointwise bootstrap confidence intervals in the *Online Appendix E*.

of 30%, increasing the share of advertising in the newspaper, increases the threshold benefit level,  $\underline{b}^r$ , and thus decreases the survival function. Hence the readers of the newspapers in our sample do not like too few advertisements. They start to get benefit after 30% market share of the advertisers. This benefit keeps increasing until 40% and after that point it starts decreasing. So the benefit that the readers get from adverts starts to decrease after a threshold point, as expected. Our results are consistent with the previous literature. In his seminal paper Rosse (1970) show that circulation demand depends positively on advertising. In a more recent study with Dutch newspapers Filistrucchi and Klein (2013) concluded the same. However, our nonparametric result further concludes that this positive effect decreases if the share of advertisements goes above 40%. This result can be explained by two effects. First, although the readers like advertisements they care more about the content of the newspaper. Second, as the number of advertisements in the newspaper increases, it causes a congestion effect as it becomes harder for the readers to find the information they are seeking from the advertisements. On the other side of the market, when we look at the estimated curves for

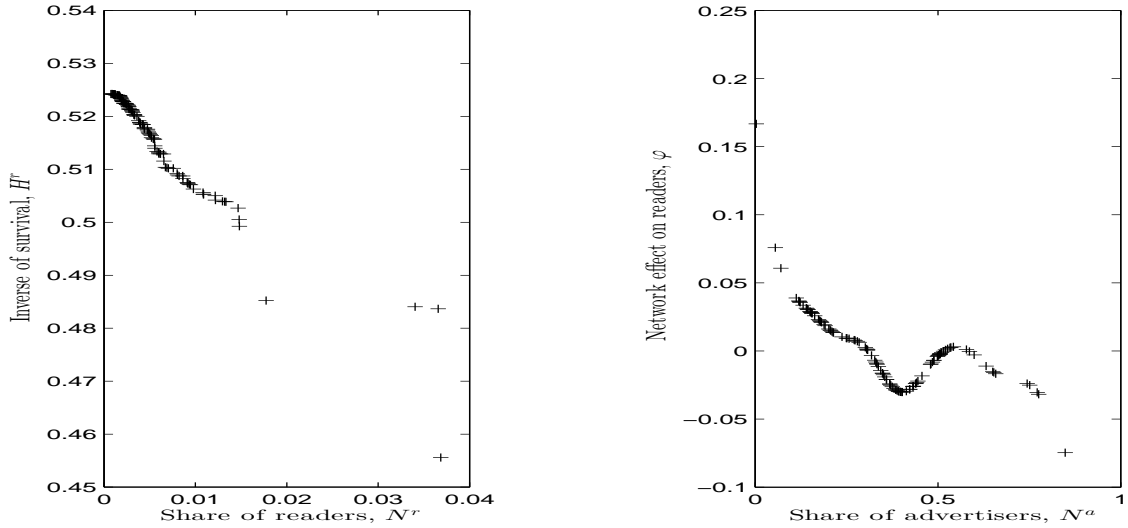


Figure 1: *Estimated functions for reader demand equation after monotonicization of  $H^r$  by rearrangement*

the advertiser demand, Figure 2, again we see that the network externality function  $\psi$  of readers changes its sign over the interval. Up to a readership share of 0.2%,  $\psi$  takes positive values meaning that it decreases the survival function. However, after 0.2%, the benefits of

advertisers are increasing with the share of readership up to 0.8%. So, we can conclude that advertisers do not benefit from advertising in a magazine with a very low level of readership. Another point worth noting is that, as for the  $\varphi$  this network effect function also follows a nonmonotonic pattern. The benefit of advertisers starts decreasing with the readership of more than 0.8%. This result may seem controversial to the results obtained in the previous literature. However, as concluded by Thompson (1989) advertisers also care about the profile of the readership. In our data set this high level of circulation may consist of profiles that are not particularly looked for by the advertisers thus their benefit start to decrease. Note that it is not only the price or the number of the pages that the newspaper chooses to maximize its profit but also the content and position of the newspaper for several dimensions, e.g. political position. Thus a high circulation can be a signal for a more moderate newspaper which is not selecting readers and leading the advertiser demand to decrease.

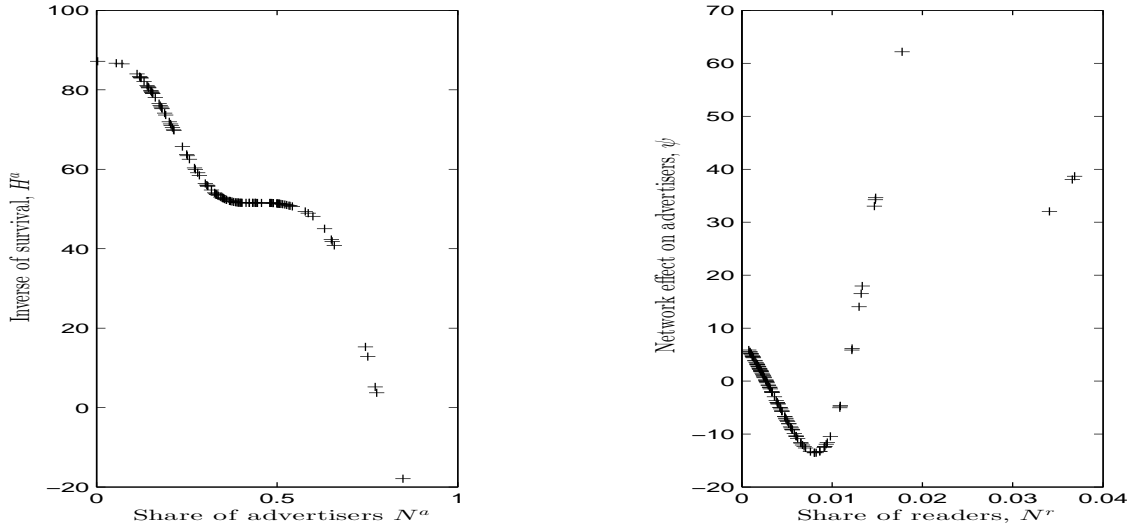


Figure 2: *Estimated functions for advertiser demand equation after monotonicization of  $H^a$  by rearrangement*

We estimate the demand equations using different set of instruments to check the robustness of our results. We present the results in *Online Appendix F*. Results of these estimations expose the same patterns such that the network effect functions are found to be nonlinear and non-monotone on both sides.

### 4.3 Parametric Estimation

In this section we estimate the model for the daily newspaper industry parametrically. The parametric specification of the network effect functions take into account the results obtained using the nonparametric specification. Using the same set of explanatory variables, we estimate the system of equations in (13) and (14) simultaneously using GMM. The results we obtain are consistent with the results of the nonparametric estimation.

#### 4.3.1 Model Specification

To do our analysis with a parametric model, we need to specify *(i)* a distribution function for the net benefits of readers and advertisers, *(ii)* a functional form for the network effects,  $\varphi(\cdot)$  and  $\psi(\cdot)$  and *(iii)* a functional form for the threshold benefit level of readers and advertisers.

First of all for the distribution of net benefits, we have chosen the log-logistic distribution function which is common in the literature of network diffusion models, see Larribeau (1993) and Feve et al. (2008). Thus, the survival function is given by:

$$S(X|m, \rho) = \frac{1}{1 + (\frac{X}{m})^\rho} \quad (25)$$

where  $m$  is the scale parameter and  $\rho$  is the shape parameter. These parameters can be estimated in advance or during the estimation of the other parameters. Larribeau (1993) assumes that  $\rho$  is constant over time while  $m$  varies and she estimates both of the parameters before estimating the demand equations. Feve et al. (2008) assume that they are constant and are equal to 1. For simplicity and without loss of generality, we also assume that both parameters are equal to 1. Secondly, we need to approximate the unknown functions  $\varphi(\cdot)$  and  $\psi(\cdot)$  by some parametric form. To do this, we use our nonparametric estimation results. We choose a third order polynomial form for reader demand equation and a second order polynomial form for advertiser demand equation. Thus:

$$\varphi(N^a) = \alpha_0 + \alpha_1 N^a + \alpha_2 (N^a)^2 + \alpha_3 (N^a)^3 \quad (26)$$

$$\psi(N^r) = \theta_0 + \theta_1 N^r + \theta_2 (N^r)^2 \quad (27)$$

Finally we need to choose a functional form for the threshold net benefit levels of the two sides. We decide to use an exponential function which will make the threshold benefit level nonnegative:

$$\underline{b}^r = f(N^a, X\beta, U) = \exp(\varphi(N^a) + X\beta + U) \quad (28)$$

equivalently for the advertisers' side:

$$\underline{b}^a = f(N^r, W\gamma, V) = \exp(\psi(N^r) + W\gamma + V) \quad (29)$$

To be consistent with the nonparametric analysis, we used the same explanatory variables, daily price and ad rate of the newspaper.

The simultaneous demand system to be estimated is:

$$\log\left(\frac{1 - N^r}{N^r}\right) = \alpha_0 + \alpha_1 N^a + \alpha_2 (N^a)^2 + \alpha_3 (N^a)^3 + \beta P^r + U \quad (30)$$

$$\log\left(\frac{1 - N^a}{N^a}\right) = \theta_0 + \theta_1 N^r + \theta_2 (N^r)^2 + \gamma P^a + V \quad (31)$$

#### 4.3.2 Estimation

Using the specification above, we estimate the demand equations, (30) and (31) simultaneously by GMM using the same instruments as before. Moreover, following Fan (2013), we use city level demographic characteristics such as percentage of females in the population, percentage of whites in the population, home ownership rate in the city, as additional instruments in both equations. We assume that these city level demographic variables are predetermined and are not correlated with the unobserved newspaper characteristics in either of the demand equations. P-value of the Hansen's J-statistic is 0.2 so we can not reject the orthogonality of instruments and the error terms. The estimation results are given in Tables (2) and (3).

All of our estimated parameters are significant at the 10% level and have the expected sign

Table 2: Estimation Results for the Reader Demand

Variable	Parameter	Estimate	Standard Error	t-stat	p-value
<i>constant</i>	$\alpha_0$	10.58	2.50	4.23	< .0001
<i>share of advertisers</i>	$\alpha_1$	-48.34	23.02	-2.10	0.0381
<i>(share of advertisers)<sup>2</sup></i>	$\alpha_2$	96.75	51.81	1.87	0.0646
<i>(share of advertisers)<sup>3</sup></i>	$\alpha_3$	-62.06	36.46	-1.70	0.0916
<i>cover price</i>	$\beta$	3.81	2.04	1.86	0.0651

Table 3: Estimation Results for the Advertiser Demand

Variable	Parameter	Estimate	Standard Error	t-stat	p-value
<i>constant</i>	$\theta_0$	1.06	0.21	5.07	< .0001
<i>share of readers</i>	$\theta_1$	-1.62	0.39	-4.10	< .0001
<i>(share of readers)<sup>2</sup></i>	$\theta_2$	3.91	1.59	2.47	0.0151
<i>ad rate</i>	$\gamma$	3.37	1.70	1.98	0.0503

For readability, estimated coefficients and standard errors of  $N^r$  ( $\theta_1$ ),  $(N^r)^2$  ( $\theta_2$ ) and  $P^a$  ( $\gamma$ ) are multiplied by  $10^{-2}$ ,  $10^{-3}$  and  $10^3$ , respectively.

in the readers' demand equation. The coefficient on price is found to be positive, meaning that an increase in the cover price will increase the threshold benefit function,  $\underline{b}^r$ , thus leading to a decrease in demand. When we examine the estimated network effect function,  $\hat{\varphi}$ , we see that our results are in line with the nonparametric estimate of  $\varphi$  in the sense that it is non-monotonic and can be approximated by a third order polynomial.

For the advertising demand equation, all of the estimated parameters, except the coefficient on  $P^a$ , are significant at the 5% level. The estimated coefficient of  $P^a$  is significant at the 10% level. It is estimated to be positive. Hence, an increase in the ad rate will increase the threshold benefit function of advertisers,  $\underline{b}^a$  leading to a decrease in advertiser demand, since the survival  $S^a$  is decreasing in its arguments. The estimation result for the  $\psi$  function is also in line with what we obtained in *Section 4.2*. Advertiser demand is increasing in the readership share until 1.8% then it starts to decrease. Moreover up to 0.8% of readership, the network effect is positive which increases threshold benefit function thus decreases the advertising demand. These numbers are found to be much smaller by the nonparametric estimation.<sup>19</sup>

To check the robustness of the nonlinear and nonmonotone network effects result, we

---

<sup>19</sup>Given the parameter estimates we also compute  $\frac{\partial S^r}{\partial N^a} \frac{\partial S^a}{\partial N^r}$  for each observation and then we take its mean. It is equal to 0.5005, hence *Assumption 7* is satisfied on average.

include more explanatory variables in each equation. The estimation results are given in *Online Appendix G*. The nonlinear and non-monotone network effects on both sides are robust to the addition of these variables.<sup>20</sup>

## 5 A Numerical Illustration

In this section we present a numerical example to demonstrate the importance of a correct specification of the network effects. To do this we estimate two parametric specifications. In the first one, the network effect function in reader demand is specified as a third order polynomial while the network effect function in advertiser demand is specified as a second order polynomial as in the model given by Equations (30) and (31). In the second one, network effects in both demand functions are specified linearly. The second specification is given by the following equations:

$$\log \left( \frac{1 - N^r}{N^r} \right) = \alpha_0 + \alpha_1 N^a + \beta P^r + U \quad (32)$$

$$\log \left( \frac{1 - N^a}{N^a} \right) = \theta_0 + \theta_1 N^r + \gamma P^a + V \quad (33)$$

We estimate the equations in each model simultaneously by GMM and we use the same set of instruments as in the previous section for both models. The estimation results of the second model are presented in Table 6 in *Online Appendix G*. All the estimated parameters, except the coefficient of daily price are significant at the 5% level. The p-value of daily price is slightly higher than 0.05. In this model as well, readers are found to be ad-lovers and an increase in the price level of the newspaper decreases the reader demand. On the other side of the market, the advertising demand increases with circulation and decreases with ad rate.<sup>21</sup>

We use the estimated parameters from both models to compute the mark-ups given in

---

<sup>20</sup>In the *Online Appendix G* we also present the graphs of  $\hat{\varphi}$  and  $\hat{\psi}$  estimated parametrically as well as we present the plots present plots which show parametric and nonparametric estimates of  $\varphi$  and  $\psi$  in one graph.

<sup>21</sup>Given the parameter estimates we also compute  $\frac{\partial S^r}{\partial N^a} \frac{\partial S^a}{\partial N^r}$  for each observation and then we take its mean. It is equal to 0.2785, hence *Assumption 7* is satisfied on average.

equations (7) and (8) at each data point. We then take the means of the computed mark-ups. First of all, in both cases we find a negative average mark-up on readers' side and a positive average mark-up on advertisers' side which is not a surprising outcome in two-sided markets. Secondly, as can be seen from equations (7) and (8), the mark-ups depend on each other, thus a specification error on one side will effect the estimated mark-ups on both sides (on top of the fact that different specification for one equation may effect the value of estimated parameters in the other equation, too.). Our results confirm these facts. The estimated mark-ups for both sides are different for different models. Estimated average mark-up on readers' side is more than 100% lower in the first model and average mark-up on advertisers' side is found to 67% higher in the first model.

This numerical illustration shows first the importance of functional specifications in empirical analysis and that an analysis which is done under a wrong specification will lead to erroneous conclusions and second how nonparametric approaches can be used as a first step to guide the researcher on functional form specification. These are especially important from a regulatory point of view. If the analysis in Argentesi and Filistrucchi (2007) is considered, where they analyze the market power of Italian newspapers, an error coming from this type of specification may reverse their results. The competitive behavior on the advertising side could have been found to be collusive if the network effects were allowed to be nonlinear.

## 6 Conclusion

This paper has developed a semi-parametric empirical model for two sided markets, more specifically for the local daily newspaper industry in the U.S. We specify the network externalities nonparametrically to be able to capture the nonlinearities and nonmonotonicities. The distribution functions of benefits of readers and advertisers are not specified and are left to be estimated, as well. The model is estimated with nonparametric IV estimation. We get two main results: First of all, the structural model is supported by the data since we obtain well behaved demand curves by nonparametric instrumental variables estimation without any restriction. Secondly, network effects on both sides are nonlinear and nonmonotone.



Using our nonparametric estimation results, we make parametric approximations for the network effect functions and re-estimate the model by GMM. We find that with a parametric specification the network externality function on the reader side is linear while on the advertiser side it is nonlinear.

This paper has many contributions. First of all, the nonparametric estimation method is a contribution as an extension to Florens and Sokullu (2014). Secondly, nonparametric specification and estimation have never been used in the empirical two sided markets literature. Finally, it proves both by nonparametric and parametric estimations that the network effects in the daily newspaper industry are neither linear nor monotone. This result indeed may have important implications for policy analysis.

The analysis uses cross-sectional data so a natural extension, nonparametric analysis with panel data is left for future work. Moreover, no dynamic model has been considered in the empirical two sided market literature. A dynamic model based on network diffusion models is a very interesting future research topic. Finally, the estimation method presented in this paper can be applied to any industry with network externalities.

## References

- Ai, C. and X. Chen (2003, November). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71(6), 1795–1843.
- Ambrus, A. and R. Argenziano (2009, February). Asymmetric Networks in Two-Sided Markets. *American Economic Journal: Microeconomics* 1(1), 17–52.
- Anderson, S. P. and S. Coate (2005). Market Provision of Broadcasting: A Welfare Analysis. *Review of Economic Studies* 72(4), 947–972.
- Andrews, D. W. (2011, May). Examples of  $l^2$ -complete and boundedly-complete distributions. Cowles Foundation Discussion Papers 1801, Cowles Foundation for Research in Economics, Yale University.
- Argentesi, E. and L. Filistrucchi (2007). Estimating market power in a two-sided market: The case of newspapers. *Journal of Applied Econometrics* 22(7), 1247–1266.
- Argentesi, E. and M. Ivaldi (2005, June). Market definition in the printed media industry: Theory and practice. CEPR Discussion Papers 5096, C.E.P.R. Discussion Papers.

- Armstrong, M. (2006). Competition in two-sided markets. *RAND Journal of Economics* 37, 668–691.
- Basu, D. (1955). *On Statistics Independent of a Sufficient Statistic*. Sankhya.
- Berry, S. T. and P. A. Haile (2014, September). Identification in Differentiated Products Markets Using Market Level Data. *Econometrica* 82(5), 1749–1797.
- Blair, R. D. and R. E. Romano (1993, April). Pricing Decisions of the Newspaper Monopolist. *Southern Economics Journal* 59(4), 721–732.
- Blundell, R. and R. Matzkin (2010, September). Conditions for the existence of control functions in nonseparable simultaneous equations models. CeMMAP working papers CWP28/10, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- Caillaud, B. and B. Jullien (2003). Chicken & egg: Competition among intermediation service providers. *RAND Journal of Economics* 34(2), 309–28.
- Carrasco, M., J.-P. Florens, and E. Renault (2007, June). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6 of *Handbook of Econometrics*, Chapter 77. Elsevier.
- Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* 96(3), 559–575.
- Chernozhukov, V., I. Fernandez-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica* 78(3), 1093–1125.
- Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011). Nonparametric instrumental regression. *Econometrica* 79(5), 1541–1565.
- Dubois, P., A. Hernandez-Perez, and M. Ivaldi (2007, 04-05). The market of academic journals: Evidence from data on french libraries. *Journal of the European Economic Association* 5(2-3), 390–399.
- D’Haultfoeuille, X. (2011, June). On the completeness condition in nonparametric instrumental problems. *Econometric Theory* 27(03), 460–471.
- Engl, H. W., M. Hanke, and A. Neubauer (1996). *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic Publications.
- Evans, D. S. and R. Schmalensee (2008). Markets with two-sided platform. *Issues in Competition Law and Policy* 1.

- Fan, Y. (2013, August). Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market. *American Economic Review* 103(5), 1598–1628.
- Ferguson, J. M. (1983, October). Daily Newspaper Advertising Rates, Local Media Cross-Ownership, Newspaper Chains, and Media Competition. *Journal of Law and Economics* 26(3), 635–654.
- Feve, F. and J.-P. Florens (2009). The practice of non parametric estimation by solving inverse problems: The example of transformation models. Technical report, Toulouse School of Economics.
- Feve, F., J.-P. Florens, F. Rodriguez, and S. Soteri (2008). Internet advertising and direct mail: Trends and analysis for the uk. Technical report, Toulouse School of Economics, Toulouse.
- Filistrucchi, L. and T. J. Klein (2013, October). Price Competition in Two-Sided Markets with Heterogeneous Consumers and Network Effects. Working Papers 13-20, NET Institute.
- Flachaire, E. (2005, June). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics & Data Analysis* 49, 361–376.
- Florens, J.-P., J. Johannes, and S. V. Bellegem (2009). Instrumental regression in partially linear models. Technical report, IDEI.
- Florens, J.-P. and S. Sokullu (2014). Nonparametric estimation of semiparametric transformation models. *mimeo*.
- George, L. (2007, October). What’s fit to print: The effect of ownership concentration on product variety in daily newspaper markets. *Information Economics and Policy* 19(3-4), 285–303.
- Hall, P. and J. L. Horowitz (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics* 32, 2904–2929.
- Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64, 103–137.
- Horowitz, J. L. (2001). The bootstrap. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5 of *Handbook of Econometrics*, Chapter 52. Elsevier.
- Horowitz, J. L. (2011, March). Applied nonparametric instrumental variables estimation. *Econometrica* 79, 347–394.

- Hu, Y. and J.-L. Shiu (2011, June). Nonparametric identification using instrumental variables: sufficient conditions for completeness. CeMMAP working papers CWP25/11, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- Kaiser, U. and M. Song (2009, March). Do media consumers really dislike advertising? an empirical assessment of the role of advertising in print media markets. *International Journal of Industrial Organization* 27(2), 292–301.
- Kaiser, U. and J. Wright (2006). Price structure in two-sided markets: Evidence from the magazine industry. *International Journal of Industrial Organization* 24(1), 1–28.
- Larribeau, S. (1993). *Microeconometrie De La Demande de Telephone*. Ph. D. thesis, Université des Sciences Sociales de Toulouse, Toulouse, France.
- Lehmann, E. and H. Scheffe (1950). *Completeness Similar Regions and Unbiased Tests Part I*. Sankhya.
- Linton, O., S. Sperlich, and I. V. Keilegom (2008). Estimation of a semiparametric transformation model. *Annals of Statistics* 36(2), 686–718.
- Newey, W. K. and J. L. Powell (2003, 09). Instrumental variable estimation of nonparametric models. *Econometrica* 71(5), 1565–1578.
- Rochet, J.-C. and J. Tirole (2003). Platform competition in two-sided markets. *Journal of the European Economic Association* 1, 990–1029.
- Rosse, J. N. (1970, March). Estimating Cost Function Parameters Without Using Cost Data: Illustrated Methodology. *Econometrica* 38(2), 256–275.
- Ryanne, B. P. and M. A. Youngson (2008). *Linear Functional Analysis*. London: Springer-Verlag.
- Rysman, M. (2004, 04). Competition between networks: A study of the market for yellow pages. *Review of Economic Studies* 71(2), 483–512.
- Thompson, R. S. (1989, March). Circulation versus Advertiser Appeal in the Newspaper Industry: An Empirical Investigation. *Journal of Industrial Economics* 37(3), 259–71.
- Weyl, E. G. (2010, September). A Price Theory of Multi-sided Platforms. *American Economic Review* 100(4), 1642–72.